

ADA033916

12  
NW

NRL Memorandum Report 3407

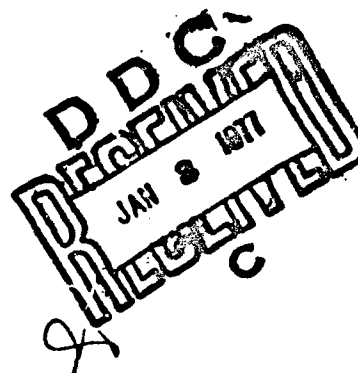
## Application of Automatic Clustering to Emitter Identification

JAMES SLAGLE AND RICHARD C. T. LEE

*Computer Science Laboratory  
Communications Sciences Division*

20000726058

November 1976



NAVAL RESEARCH LABORATORY  
Washington, D.C.

Approved for public release; distribution unlimited.

Reproduced From  
Best Available Copy

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER NRL Memorandum Report 3407	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) APPLICATION OF AUTOMATIC CLUSTERING TO EMITTER IDENTIFICATION	5. TYPE OF REPORT & PERIOD COVERED As interim report on a continuing NRL problem.	
7. AUTHOR(s) James Slagle and Richard C.T. Lee	8. CONTRACT OR GRANT NUMBER(s)	
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Research Laboratory Washington, D.C. 20375	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NRL Problem B02-23	
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Electronic Systems Command Washington, D.C. 20360	12. REPORT DATE November 1976	
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	13. NUMBER OF PAGES 12	
	15. SECURITY CLASS. (of this report) UNCLASSIFIED	
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE	
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from (16))		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Clustering Unsupervised learning Pattern recognition Nearest neighbor Algorithm Feature selection algorithm		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) The goal of clustering is the partitioning of a given set of objects into subsets called clusters in such a way that the objects in a cluster are similar to one another and that objects in different clusters are dissimilar. Clustering may help in getting a more or less direct understanding of the relationships among the objects, and it may be useful as a first step in pattern recognition. Some possible applications are automatic phoneme recognition, data base management systems, personnel classification, detection of errors in files and computer security. (Continues)		

DD FORM 1473

JAN 73

EDITION OF 1 NOV 65 IS OBSOLETE  
S/N 010-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

251 150

for

## 20. Abstract (Continued)

Several clustering methods were applied to data sets of practical importance. Automatic pattern recognition using the  $k$  nearest neighbors was applied. An efficient method for selecting a good subset from the full set of 44 features was tried. In all cases, the results were good.

CLASSIFIED BY	DATE	BY
EXEMPTED FROM AUTOMATIC DOWNGRADING AND DECLASSIFICATION		
EXEMPTION CODE	EXEMPTION AUTHORITY	EXEMPTION DATE
1	2	3
4	5	6
7	8	9
10	11	12
13	14	15
16	17	18
19	20	21
22	23	24
25	26	27
28	29	30
31	32	33
34	35	36
37	38	39
40	41	42
43	44	45
46	47	48
49	50	51
52	53	54
55	56	57
58	59	60
61	62	63
64	65	66
67	68	69
70	71	72
73	74	75
76	77	78
79	80	81
82	83	84
85	86	87
88	89	90
91	92	93
94	95	96
97	98	99
100	101	102
103	104	105
106	107	108
109	110	111
112	113	114
115	116	117
118	119	120
121	122	123
124	125	126
127	128	129
130	131	132
133	134	135
136	137	138
139	140	141
142	143	144
145	146	147
148	149	150
151	152	153
154	155	156
157	158	159
160	161	162
163	164	165
166	167	168
169	170	171
172	173	174
175	176	177
178	179	180
181	182	183
184	185	186
187	188	189
190	191	192
193	194	195
196	197	198
199	200	201
202	203	204
205	206	207
208	209	210
211	212	213
214	215	216
217	218	219
220	221	222
223	224	225
226	227	228
229	230	231
232	233	234
235	236	237
238	239	240
241	242	243
244	245	246
247	248	249
250	251	252
253	254	255
256	257	258
259	260	261
262	263	264
265	266	267
268	269	270
271	272	273
274	275	276
277	278	279
280	281	282
283	284	285
286	287	288
289	290	291
292	293	294
295	296	297
298	299	300
301	302	303
304	305	306
307	308	309
310	311	312
313	314	315
316	317	318
319	320	321
322	323	324
325	326	327
328	329	330
331	332	333
334	335	336
337	338	339
340	341	342
343	344	345
346	347	348
349	350	351
352	353	354
355	356	357
358	359	360
361	362	363
364	365	366
367	368	369
370	371	372
373	374	375
376	377	378
379	380	381
382	383	384
385	386	387
388	389	390
391	392	393
394	395	396
397	398	399
400	401	402
403	404	405
406	407	408
409	410	411
412	413	414
415	416	417
418	419	420
421	422	423
424	425	426
427	428	429
430	431	432
433	434	435
436	437	438
439	440	441
442	443	444
445	446	447
448	449	450
451	452	453
454	455	456
457	458	459
460	461	462
463	464	465
466	467	468
469	470	471
472	473	474
475	476	477
478	479	480
481	482	483
484	485	486
487	488	489
490	491	492
493	494	495
496	497	498
499	500	501
502	503	504
505	506	507
508	509	510
511	512	513
514	515	516
517	518	519
520	521	522
523	524	525
526	527	528
529	530	531
532	533	534
535	536	537
538	539	540
541	542	543
544	545	546
547	548	549
550	551	552
553	554	555
556	557	558
559	560	561
562	563	564
565	566	567
568	569	570
571	572	573
574	575	576
577	578	579
580	581	582
583	584	585
586	587	588
589	590	591
592	593	594
595	596	597
598	599	600
601	602	603
604	605	606
607	608	609
610	611	612
613	614	615
616	617	618
619	620	621
622	623	624
625	626	627
628	629	630
631	632	633
634	635	636
637	638	639
640	641	642
643	644	645
646	647	648
649	650	651
652	653	654
655	656	657
658	659	660
661	662	663
664	665	666
667	668	669
670	671	672
673	674	675
676	677	678
679	680	681
682	683	684
685	686	687
688	689	690
691	692	693
694	695	696
697	698	699
700	701	702
703	704	705
706	707	708
709	710	711
712	713	714
715	716	717
718	719	720
721	722	723
724	725	726
727	728	729
730	731	732
733	734	735
736	737	738
739	740	741
742	743	744
745	746	747
748	749	750
751	752	753
754	755	756
757	758	759
760	761	762
763	764	765
766	767	768
769	770	771
772	773	774
775	776	777
778	779	780
781	782	783
784	785	786
787	788	789
790	791	792
793	794	795
796	797	798
799	800	801
802	803	804
805	806	807
808	809	810
811	812	813
814	815	816
817	818	819
820	821	822
823	824	825
826	827	828
829	830	831
832	833	834
835	836	837
838	839	840
841	842	843
844	845	846
847	848	849
850	851	852
853	854	855
856	857	858
859	860	861
862	863	864
865	866	867
868	869	870
871	872	873
874	875	876
877	878	879
880	881	882
883	884	885
886	887	888
889	890	891
892	893	894
895	896	897
898	899	900
901	902	903
904	905	906
907	908	909
910	911	912
913	914	915
916	917	918
919	920	921
922	923	924
925	926	927
928	929	930
931	932	933
934	935	936
937	938	939
940	941	942
943	944	945
946	947	948
949	950	951
952	953	954
955	956	957
958	959	960
961	962	963
964	965	966
967	968	969
970	971	972
973	974	975
976	977	978
979	980	981
982	983	984
985	986	987
988	989	990
991	992	993
994	995	996
997	998	999
1000	1001	1002

## CONTENTS

1. INTRODUCTION .....	1
2. CLUSTERING THE EMITTER IDENTIFICATION DATA .....	1
3. TWO CLUSTERING METHODS .....	3
4. FUTURE WORK .....	8
REFERENCES .....	9

## APPLICATION OF AUTOMATIC CLUSTERING TO EMITTER IDENTIFICATION

### 1. INTRODUCTION

Research in and applications of artificial intelligence [4] are the main thrusts of the Computer Science Laboratory, Communications Sciences Division, Naval Research Laboratory. At present, we are working on phoneme recognition for a low band width speech communication system, a computer-controlled manipulator, and intelligent data base management system, as well as automatic clustering and its application to emitter identification. In this paper, we shall focus on the automatic clustering work. The goal of clustering [1,2,3,5,6] is the partitioning of a given set of objects into subsets called clusters in such a way that the objects in a cluster are similar to one another and that objects in different clusters are dissimilar. Two objects may be considered similar if, for example, the euclidean distance between them in the measurement (feature) space is small.

Clustering has two main purposes. First, clustering may help in getting a more or less direct understanding of the relationships among the objects. Second, clustering may be useful as a first step in pattern recognition. In pattern recognition (unlike clustering), each presented object must be labeled with its class membership. After clustering (with labels ignored), one can "look at" the data to estimate whether pattern recognition will be easy or difficult, whether a given class should be combined with another class or divided into two or more classes, etc.

Since clustering is quite general, it has many applications. Later we shall see that clustering can be applied to emitter identification. It can be applied to automatic phoneme recognition and personnel classification. It is often very important to find a cluster having only one member. Such a cluster may be a mistake in the data base or represent a very unusual object. An unusual use of a computer system may be an attempted security penetration.

### 2. CLUSTERING THE EMITTER IDENTIFICATION DATA

Each row in Table 1 is a sample (object) representing 18 measurements of an emission. Our sponsor gave us 18 samples. Without knowing anything else about the data, we applied several clustering techniques, some of which we had developed [2,5,6]. We obtained the following four clusters:

- (1) 1, 2
- (2) 3, 5, 6, 7, 9, 10, 11
- (3) 4, 17, 18
- (4) 8, 12, 13, 14, 15, 16

This means that samples (rows) 1 and 2 are in cluster (1), etc.

Note: Manuscript submitted October 29, 1976.

Table 1. Original Data for Emitter Identification

PARAMETERS																	
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	-.10	-.07	-1.26	-.76	.74	2.47	-.46	-.44	-.97	-.36	-.36	-.46	-.44	-.97	-.50	0.00	.14
2	-.33	-.40	-1.02	-.76	.12	1.95	-.44	-.27	-1.80	-1.80	-.97	-.44	-.27	-1.80	-1.00	-.53	-.62
3	1.01	1.23	1.18	.85	-1.17	.26	.85	1.70	1.60	1.60	1.60	.85	1.70	1.60	.50	.36	.38
4	-.99	-1.05	1.31	.85	-.37	.75	-.46	-.56	.32	.32	.32	-.46	-.56	.32	-.33	-.79	.83
5	-1.24	1.23	.71	.85	-.71	.22	-.15	-1.05	1.00	1.00	1.00	-.15	-1.05	1.00	.83	.64	.31
6	1.02	.90	.35	.44	-.61	.49	-.80	.19	-.36	-.36	-.36	-.80	.19	-.36	2.50	3.16	.90
7	1.57	1.55	.65	.85	-.60	.21	-.18	-.98	1.03	1.03	1.03	-.18	-.98	1.03	-1.50	-.64	.13
8	-.49	.72	-.96	-1.56	.27	-.65	-.40	.28	-1.58	-1.58	-1.58	-.40	.28	-1.58	-.33	0.00	.23
9	.13	-.40	.61	.45	-1.43	.91	-.62	-1.34	1.23	1.23	1.23	-.62	-1.34	1.23	.67	.64	.33
10	1.48	1.55	.71	.85	3.19	-.70	.82	.42	1.16	1.16	1.16	.82	.42	1.16	.33	.71	.36
11	.53	.58	.28	.45	-.40	.50	.78	.27	.36	.36	.36	.78	.27	.36	1.33	.64	.43
12	.25	-.40	-.97	-1.16	-.33	-.95	-.39	.20	-.97	-.97	-.97	-.39	.20	-.97	1.67	-.64	-.52
13	-.61	-.72	1.05	-1.16	.48	-.98	-.37	-.88	.58	.58	.58	-.37	-.88	.58	-.33	-1.43	-.62
14	1.48	-1.37	-1.13	-.76	.66	-.69	3.43	-.19	-.39	-.39	-.39	3.43	-.19	-.39	.17	.36	.27
15	-.03	.25	.94	-.76	.45	-.24	-.37	-.14	-.29	-.29	-.29	-.37	-.14	-.29	-1.50	.64	.27
16	-1.85	-1.70	-.08	-1.16	.50	-.23	-.37	1.13	1.03	1.03	1.03	-.37	1.13	1.03	-.50	-.29	.18
17	-.61	-.40	1.47	1.65	-.46	-.67	-.54	.21	-1.52	-1.52	-1.52	-.54	.21	-1.52	.17	.21	.29
18	-.94	-.72	1.01	1.25	.17	-1.13	-.51	2.89	.55	.55	.55	-.51	2.89	.55	-.17	-.50	.11

Our sponsor then told us that the "real" clusters should be as follows:

- (1') 5, 6, 10, 11
- (2') 4, 17, 18
- (3') 3, 7, 9
- (4') 1, 2, 8, 12, 13, 14, 15, 16

Thus, the results were quite good. Careful analysis of the data revealed the following:

- (a) If cluster (1) has to be combined with any other cluster, it should be combined with cluster (4). However, cluster (1) is definitely different from cluster (4).
- (b) From only the given data, one really cannot say that cluster (1') and cluster (3') should be two clusters rather than being combined into one cluster.

Our sponsor agreed with both of these statements.

We were then supplied with 12 new samples (rows) with some missing values. (See Table 2). After using various clustering methods, we concluded that A, B, C, D, E, and F belong to cluster (1'), that G, H, and I belong to (2'), and that J, K, and L belong to either (1') or (3'). Then our sponsor told us that the first two conclusions are absolutely correct. He said that samples J, K, and L belong to cluster (3'). We could say only that they belong to either (1') or (3'), because (1') and (3') are so close to each other.

### 3. TWO CLUSTERING METHODS

Two of the six clustering methods we used are data reorganization [5] and two principal components [1,3]. The other four are minimal spanning trees [1,3], non-linear mapping [6], and two versions of a triangulation method [2]. We do not describe these here, since they are fairly complicated to explain and are described elsewhere. In data reorganization, rows (and sometimes columns as here) are permuted to put similar rows (and columns) together. (Two n-tuples are similar to the extent that the n-dimensional distance between them is small.) The result for the 18 samples is shown in Table 3. After this reorganization was done, the sponsor told us the classes. Table 3 is perfect in the sense that it could be broken between rows into four parts, each exactly corresponding to a given class. We also used the reorganization by rows for automatically obtaining the clustering hierarchy shown in Fig. 1. Table 3 is broken into two tables by dividing between the rows whose distance apart is largest. These tables are divided further, etc. Partly on this hierarchy, we obtained the clusters (1) through (4) given earlier.

Table 2. Emitter Identification Data With Missing Values

PARAMETERS														
A	1.65	1.55	.92	.42	.85	-.23	1.62	.78	—	—	.33	.21	.28	
B	1.70	1.55	.66	.44	—	—	—	-.16	—	—	—	—	—	
C	1.19	1.23	.65	.44	-.81	.30	1.39	.76	-.82	1.23	1.50	.86	—	
D	1.15	1.23	.91	1.25	-.81	-.40	—	-.12	—	—	—	—	—	
E	1.65	1.55	.68	.85	-1.28	.35	—	.82	—	—	—	—	—	
F	1.43	1.23	.59	.45	-1.19	.40	.43	-.20	2.32	—	—	—	—	
G	-.22	-.40	1.60	1.25	0.58	-.82	—	-.62	—	—	—	—	—	
H	-1.37	-1.37	1.29	1.65	.26	-.70	—	.69	—	—	—	—	—	
I	-.18	-.40	1.71	1.25	—	—	—	.70	—	—	—	—	—	
J	1.32	1.23	1.08	1.25	-.78	.49	—	-.15	—	—	—	—	—	
K	1.41	.90	.99	.85	—	—	—	.65	—	—	—	—	—	
L	.46	.25	.93	.45	—	—	—	-.20	—	—	—	—	—	
EMISSIONS														

Table 3. Data Reorganized by Row and Column

	14	18	17	15	3	9	4	15	7	8	1	2	12
1	0.44	0.14	0.00	-0.50	-1.26	-1.04	-0.76	-0.97	-1.06	0.00	-0.10	-0.07	1.22
2	0.27	-0.62	-0.43	-1.00	-1.02	-1.04	-0.76	-1.80	-0.51	0.00	-0.33	-0.40	0.97
12	0.20	-0.52	-0.64	-1.67	-0.97	-1.04	-1.16	-0.97	-0.68	0.00	0.25	-0.40	-1.30
8	0.28	0.23	0.00	-0.33	-0.96	-1.04	-1.56	-1.58	-0.64	-0.97	-0.49	0.72	-1.07
14	-0.19	0.26	0.36	0.17	-1.13	-1.04	-0.76	-0.84	-0.58	-0.97	-1.48	-1.37	-1.24
16	1.13	0.18	-0.29	-0.50	-0.97	-1.04	-1.16	1.03	0.45	-1.94	-1.85	-1.70	-0.88
13	-0.88	-0.62	-1.43	-0.33	1.05	-1.04	-1.16	0.58	0.13	0.00	-0.61	-0.72	-1.05
15	-0.15	0.27	0.64	-1.50	0.94	-1.04	-0.76	-0.29	-0.11	0.00	0.33	0.25	-0.51
10	0.42	0.35	0.71	0.33	0.71	0.57	0.85	1.16	1.52	0.97	1.48	1.55	1.01
6	0.19	0.90	<u>3.36</u>	2.50	0.35	0.57	0.44	-0.36	1.38	0.97	1.02	0.90	1.04
11	0.37	0.43	0.64	1.33	0.29	0.57	0.45	0.36	1.30	0.97	0.53	0.58	0.98
5	1.05	0.31	0.64	0.83	0.71	0.57	0.85	1.00	1.16	0.97	1.24	1.23	1.01
7	-0.98	0.13	-0.64	-1.50	0.65	0.57	0.85	1.03	1.36	1.94	1.57	1.55	1.02
3	-1.70	0.38	0.36	0.50	1.18	1.11	0.85	1.16	0.78	0.97	1.01	1.23	0.81
9	-1.34	0.33	0.64	0.67	0.61	0.57	0.45	1.23	-0.63	0.00	0.13	-0.40	0.44
4	0.45	0.83	-0.79	-0.33	1.31	1.11	0.85	0.32	-1.66	-0.97	-0.99	-1.05	-1.01
17	0.21	0.29	0.21	0.17	1.47	1.65	1.65	-1.52	-1.08	-0.97	-0.61	-0.40	-0.65
18	<u>2.89</u>	0.11	-0.50	-0.17	1.01	1.11	1.25	0.55	-0.22	-0.97	-0.94	-0.72	-0.80

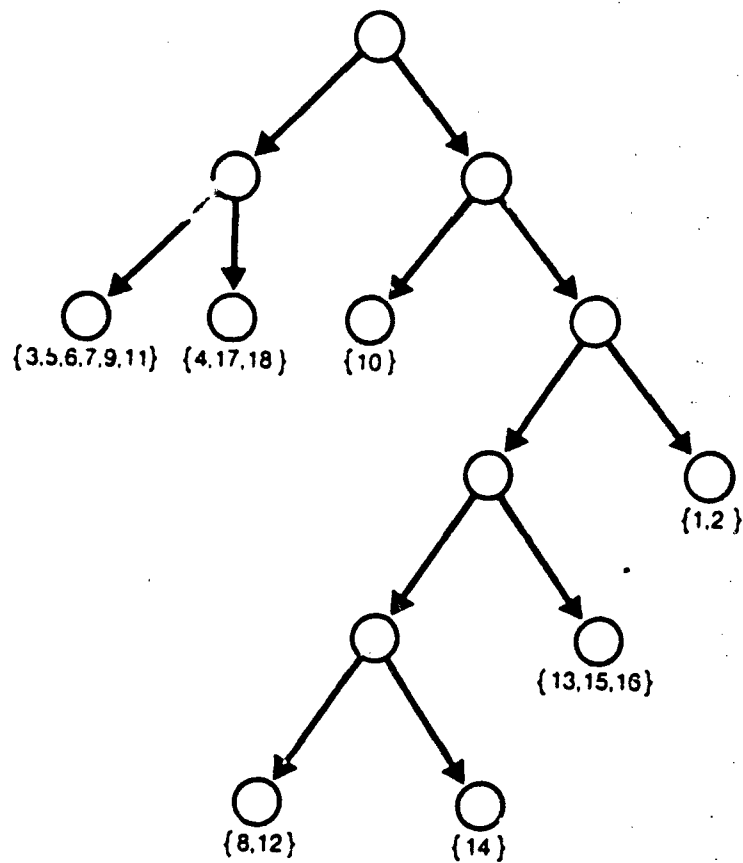


Fig. 1 — Clustering hierarchy from data reorganization by row

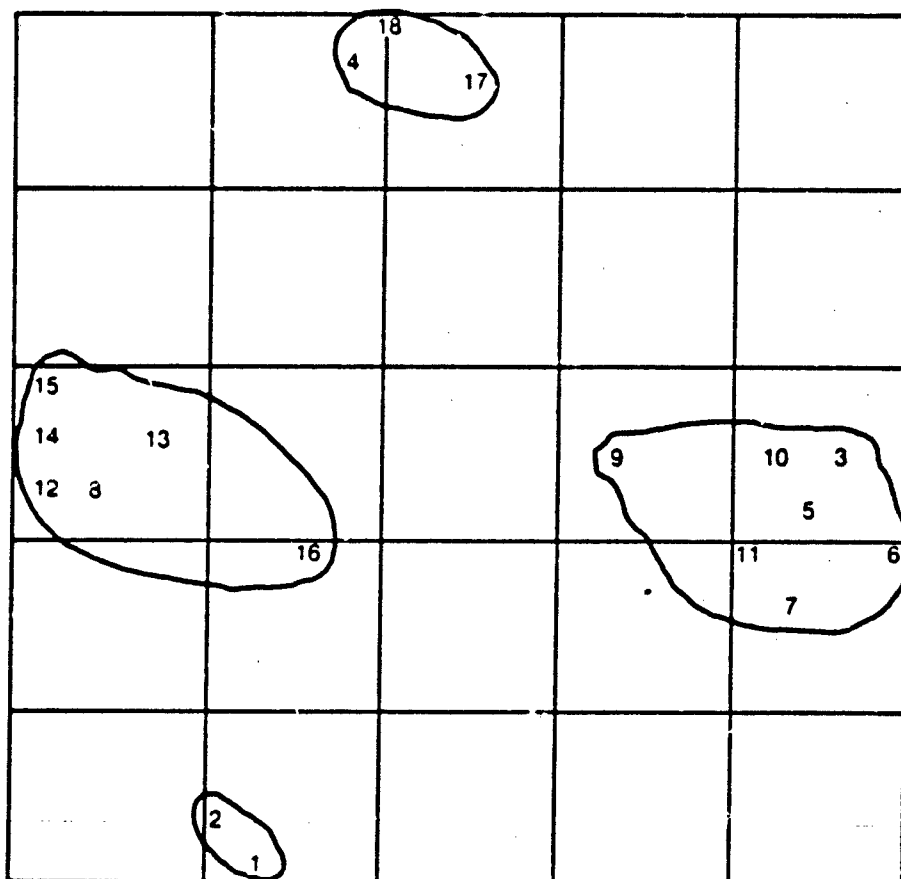


Fig. 2 — Principal components

In the method of two principal components, the data is projected from an 18 dimensional space onto the "best" plane (two dimensional space). The result for our data is shown in Fig. 2. Each of the two axes is a linear combination of the original 18 features. Again we see that the results are good. Partly based on Fig. 2, we obtained clusters (1) through (4) given earlier.

#### 4. FUTURE WORK

Our sponsor will give us a much larger set of data, and we are looking forward to analyzing it. We hope that the results will continue to be favorable. We shall also be applying clustering techniques to automatic phoneme recognition and intelligent data base management systems.

#### REFERENCES

- [1] Duda, R. and Hart, P.: "Pattern Classification and Scene Analysis". Wiley Interscience, 1973.
- [2] Lee, R. C. T., Slagle, J., and Blum, H.: "A Triangulation Method for the Sequential Mapping of Points from n-Space to 2-Space". (To be published as correspondence in IEEE Transactions on Computers.)
- [3] Misel, W.: "Computer Oriented Approaches to Pattern Recognition". Academic Press, 1972.
- [4] Slagle, J.: "Artificial Intelligence: The Heuristic Programming Approach." McGraw-Hill, New York, 1971. Also available in German and Japanese translations.
- [5] Slagle, J., Chang, C.L., and Heller, S.: A Clustering and Data Reorganizing Algorithm, IEEE Trans. on Systems, Man, and Cybernetics, Vol. SMC-5, (Jan. 1975) pp. 125-128.
- [6] Slagle, J., Chang, C.L., and Lee, R.C.T.: Experiments with some Cluster Analysis Algorithms", Pattern Recognition, Vol. 6, 1974, pp. 181-187.